图书情报工作
第62卷 第23期 2018年12月

政府网站信息资源知识元模型与可视化表征研究*

王萍[1] 王美月[1] 王益成[1] 黄新平[2]
[1] 吉林大学管理学院 长春 130022 [2]清华大学公共管理学院 北京 100084

摘要：[**目的/意义**]利用知识元模型理论研究政府网站知识服务效果的优化路径，辅以可视化表征技术，以降低大数据环境下政务用户信息获取的操作负载和知识加工的认知负荷。[**方法/过程**]依据相关知识元模型研究推理出符合政府网站信息资源属性特征的六元组知识元表示方法和四元组知识元本体结构，采用TextRank与HDP算法分别抽取政府网站信息资源关键词和主题词，并由领域专家根据抽取结果确定知识元，构建包含知识元本体库生成和可视化知识服务的政府网站信息资源领域知识元可视化表征模型。[**结果/结论**]通过政府网站发布的共享单车实例检验知识元可视化表征模型的有效性和可行性，为实现政府网站粗粒度信息服务转向以知识元为单位的细粒度知识服务范式开辟了新的研究思路，可视化知识服务模式增强了政务信息导航的结构化和用户解读领域文本语义的效果。

关键词：政府网站信息资源 知识元 本体 可视化表征
分类号：G250
DOI：10.13266/j.issn.0252-3116.2018.23.002

1 引言

我国情报学家徐如镜指出以文献为控制单位的知识资源开发与服务已无法满足大数据环境下人们知识粒度的需求，突破以"文献"为单元，将知识控制单位深化为"知识元"(知识的最小单元)，深度揭示知识内在关联，将极大促进用户对知识的有效获取、理解与利用，实现知识创新及增值[1]。政府网站作为政府公共服务的主要渠道，为用户提供全生命周期的各类公开信息与服务，以此来推进政务信息共享、网上便民服务、宣传政府各项建设成就及重大决策与活动等。随着互联网与移动技术的发展，电子政务以其移动性、互通性、价格低廉等优势迅速普及并应用在各级政府机构，据中国互联网络信息中心(CNNIC)发布的《第41次中国互联网络发展状况统计报告》[2]显示，2017年我国在线政务服务用户规模已达到4.85亿，截止到2017年12月我国共有"GOV.CN"域名47 941个，分布在31个省、自治区和直辖市，政府门户网站已成为公众获取政务信息资源的最主要渠道。然而，由于政府机构的层级性，各级部门缺乏统一的标准规范，导致应用系统自成体系，发布的海量、分布式的信息资源未能有效地组织与管理，降低了公众获得有效信息的效率，严重影响政府公众服务效能。因此迫切需要构建一个科学有效的知识组织服务体系，保障公众知识的有效获取与利用。本文基于知识元模型理论，构建政府网站信息资源知识元模型，并采用可视化表征形式向用户呈现结构化的内容，降低用户知识获取过程中的认知障碍，从而推动政府网站信息资源建设与高效流转。

2 研究现状

2.1 政府网站信息资源组织相关研究

随着我国信息技术的快速发展，电子政务在提高政府行政、服务与管理方面发挥着重要的作用。大数据环境下政府网站信息资源呈现信息量大、更新速度快、内容涉及面广、存储分散等特点，如何对其进行有效组织与管理以提升政府网站知识服务能力是该领域研究的重要议题。目前相关的理论与实践研究已取得了一定成果，研究主要聚焦在领域本体、云计算、主题地图以及关联数据等方面。B. Bouguettaya[3]等以分布

* 本文系国家自然科学基金应急管理项目"政府网站信息资源多维语义知识融合研究"(项目编号:71740015)研究成果之一。
作者简介：王萍(ORCID:0000-0002-0308-7362)，教授，博士，博士生导师；王美月(ORCID:0000-0002-3572-3511)，博士研究生，通讯作者，E-mail:wangmeiyue328@163.com；王益成(ORCID:0000-0001-6242-1765)，博士研究生；黄新平，(ORCID:0000-0003-2658-2195)，助理研究员，博士。
收稿日期:2018-04-02 修回日期:2018-08-10 本文起止页码:14-21 本文责任编辑:易飞

式本体模型为基础,开发了一个能够对政府数据和电子服务组织与动态管理的原型系统。G. Prokopiadou[4]等学者在分析政务信息资源特征基础上提出了一种面向知识本体的政务公共信息资源管理与传播方法。我国学者高洁与李佳培[5]、耿瑞利[6]等利用政府主题词构建领域本体的方法组织政务信息资源。邓峰等[7]构建基于云计算的政务信息资源融合平台框架,详细阐述了构建资源池、云平台以及服务实施3个流程的方法,并针对平台存在的问题提出了相应的解决方案。吕元智[8]构建了由资源层、管理中间件层和服务层组成的电子政务信息资源共享云计算实现理论模型,推动了云计算技术在政务信息资源管理中的应用。张玉涛、夏立新[9]构建了基于主题图的电子政务信息资源整合模型,并展示了 Metamorposis 主题图环境下资源整合的流程。此外,应用关联数据[10]实现语义层面的政务信息资源组织方法也为该领域研究提供了一个新的思路。

2.2 知识元相关研究

信息技术、社会环境及人们认知需求变化所引发的知识组织方式变革是大数据时代的必然结果,由此也扩展了基于知识元的知识组织、知识挖掘相关研究和应用领域,目前已有的研究已深入到图书馆学、情报与档案管理、计算机科学、教育学、医学等多个学科。温有奎[11]提出知识元与向导信息的导航变化理论和方法,该理论明确了知识元在知识结构中的属性特征,知识元导航链接功能以及知识元与向导信息的组合增值变换规律,为实现以知识元为单位的知识组织、检索与集成奠定了理论基础。姜永常等[12]以知识元为基元与知识链接共同构建知识网络和语义网,并阐述了知识元在知识组织中的知识发现、创新、挖掘与评价的服务功能。王宇与李秀秀[13]以知识元作为文献知识组织的最小单位,通过知识元表示、抽取分类、链接以及建立知识元库实现期刊文献的知识组织与检索。毕崇武等[14]分析了数字图书馆用户多粒度知识服务的需求,构建了基于知识元的数字图书馆组织服务模式,研究表明,该方法能有效的为用户提供分层的、不同粒度的知识资源,满足不同用户知识需求。此外,在隐性知识发现[15]、风险预测与评估[16]等方面知识元的组织方式也发挥着重要的作用。

综上所述,目前政府网站信息资源组织的研究取得一定进展,但在知识检索、知识组织及利用方面仍是以文献为控制单位的粗粒度信息服务模式。本文借鉴其他领域知识组织的成功经验,结合政府网站知识结构的特征,构建政府网站信息资源知识元模型,并从认知心理学的视角出发,利用可视化技术以形象的符号体系反映政务信息资源,从而加强用户对文本语义的理解[17],降低用户认知负荷,提高用户知识获取效率。

3 政府网站信息资源知识元模型构建

政府网站信息资源以数据形式描述和存储,而计算机无法直接识别自然语言中的知识,需要通过构建领域知识模型对知识进行符号化表征。采用知识元表示政府网站信息资源领域知识,并以知识元之间的相互关系描述领域知识结构,将有助于知识组织从文献层面深入到知识元层面,实现知识的高效组织与管理。

3.1 政府网站信息资源领域知识元表示

政府网站信息资源具有共享性、可复制性、再生性以及与载体不可分性等[18]特点,知识元是构成领域知识的最小单元,统一的知识元表示规范能有效地整合资源,促进知识的有效获取、存储与利用。随着知识管理与知识服务领域研究的不断深入,学者们依据不同领域的知识属性,提出了二元组、三元组、五元组及七元组的知识元模型,如表1所示:

表1 知识元表示方法

作者	知识元模型	知识元属性
周宁等[19]	二元组	名、值
温有奎等[20]	三元组	对象名、属性集、状态集
高国伟等[21]	三元组	概念、关系、问题
毕经元等[22]	五元组	链接、来源、名称、作用、内容
余洋[23]	七元组	知识元标志号、知识名称、关键字集、简要说明、知识类别、知识级别、知识地址

本文综合上述知识元模型,结合我国《政府信息公开目录系统实施指引(试行)》[24]发布的核心元数据标准提出六元组知识元表示,即:

$$KE = <I, T, K, D, C, O>$$

其中,KE 为知识元,I、T、K、D、C、O 分别为索引号、名称、关键词、内容描述、类别、地址,具体描述如下所示:

索引号(identifier, I):公开信息的标识符,主要用于跨地区、跨部门政府信息公开和共享的计算应用系统中,是每条信息唯一的标识。

名称(title, T):政府公开信息,是对发布内容的高度概括。

关键词(keywords, K):指反映政府公开信息内容特点的词语,包括主题关键词和位置关键词。

内容描述(description, D):政府公开信息内容的概

述,包括政策法规、政府公告、工作动态等。该属性是用户获取精准知识内容,提高知识检索与利用效率的前提。

类别(category,C):指政府公开信息所属类别的标示,包括类目名称和分类代码。政府信息资源分类方式包括:主题分类、机构分类、体裁分类以及服务对象分类。

地址(online,O):指政府公开信息网上统一资源标识符,用来标识信息来源所在载体,用户通过访问网址可获取完整的信息内容。

上述知识元全面涵盖了单篇政府文献的主要信息,依照上述表示方法在文本信息中抽取全部相关属性内容,有利于深层次的知识挖掘与知识元本体结构的构建。

3.2 政府网站信息资源知识元本体结构

Y. Yao 等[25]指出语义知识检索系统不是通过标引来组织数据和文献,而是通过知识元的连接来组织,政府网站信息资源知识元本体的建立能够揭示知识元及其相互之间复杂的语义关系,明确知识元属性与结构,是实现语义知识检索的基础。知识元本体库由关系特征库、元数据信息库、知识元语义图构成。其中关系特征库是存储知识元本体的各种关系,包括并列关系、关联关系、包含关系等,是知识元交流与推理的基础;元数据信息库是存储知识元本体数据类型;知识元语义图由内部的知识元实体及外部的知识元间关系构成,可将知识元的结构描述为四元组[4]:

$$k = <c,\ p, m, r>$$

其中,k 表示知识元本体元素,c 表示某个领域概念,p,m 分别是概念 c 上的一组属性和方法,r 是建立在 c 上与其他概念的一组关系。根据上述对知识元本体结构的描述,若知识元本体由 n 个本体元素组成,则概念集 $C = \{c_1, c_2, c_3, \cdots, c_n\}$、属性集 $P = \{p_1, p_2, p_3, \cdots, p_n\}$、方法集 $M = \{m_1, m_2, m_3, \cdots, m_n\}$ 以及建立在 C 上的关系集 $R = \{r_1, r_2, r_3 \cdots, r_n\}$ 将共同组成一个有向图 G,$G = \{X; E\}$,X 为图的点集,E 为边集,X 的值域为概念集 C,E 的值域为关系集 R。见图 1。

鉴于政府网站信息资源动态更新与增长,知识元本体结构与知识元关系也随之改变,受本体自动化更新技术的局限,为了保证领域知识元本体演化的效率与准确性,进化过程中应在领域专家的干预下操作完成。知识元本体的引入,明确了知识元所在领域本体相关概念的位置及知识元间的组合链接,是政府网站信息资源实现知识元可视化的关键。

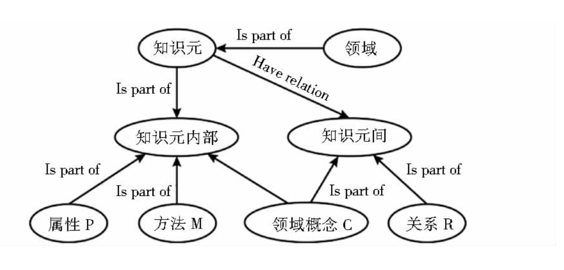


图 1 知识元本体语义图[1]

4 政府网站信息资源知识元抽取与可视化表征

4.1 政府网站信息资源知识元抽取

知识元抽取是政府网站信息资源知识存储、管理及可视化的前提,以政府信息资源关键词为基础,通过关键词识别,判断其是否包含知识元。由于政府网站信息资源涉及单个领域知识语料库较单一,抽取结果会出现关键词冗余度高、主题难以聚类等问题,本文采用关键词与主题模型相结合的方法,比对分析抽取后的关键词,获取更全面的领域知识元。

4.1.1 TextRank 关键词抽取 关键词抽取的主要目的在于从单篇文档中自动提取能够表达整篇文档主题内容的词语,最典型的关键词抽取算法是基于词汇共现图的 TextRank 算法[26],TextRank 算法根据单篇文献词汇共现图中一个词汇节点越重要,与其连接的词汇节点就越重要这一假设,评估词汇共现图中每一个词汇节点的权重,权重越大的词汇则越可能表达这篇文章的主旨,因此本文关键词的确立取决于权重的大小排序。关键词抽取的基本流程如图 2 所示:

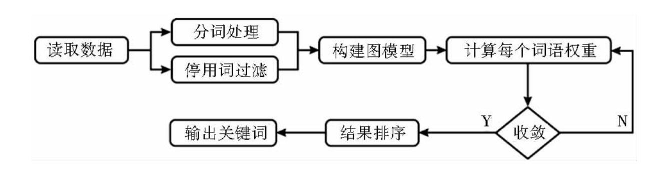


图 2 TextRank 算法实现流程

(1)将文本 T 分割成若干句子,即:$T(S_1, S_2, \cdots, S_n)$。

(2)对逐个句子 $Si \in T$ 进行分词、停用词过滤处理,从而获得句子集与单词集。

(3)构建候选关键词共现图 $G = (V, E)$,其中 V 为节点集,由候选关键词组成;E 为边集,由具有共现关系的两点构成的边组成。设窗口大小为 K,假设一个句子由词 $w_1, w_2, w_3 \cdots, w_n$ 组成,$[w_1, w_2, \cdots, w_k]$,$[w_2, w_3, \cdots, w_{k+1}]$,…,$[w_{n-k+1}, w_{n-k+2}, \cdots, w_n]$ 等各为一个窗口,在每个窗口中的任意两个词汇节点之间存在边,

则构成一个无向无权图。

(4)对每一单词，计算权重 S(Vi)，使用公式(1)迭代计算文本中任一词语 Vi 的权值，式中 d 为调节系数，一般取 0.85，S(Vi)初始值设为 1，当图中任意一个节点权重的迭代误差值小于 0.0001 时收敛，停止迭代。

$$S(v_i) = (1-d) + d \times \sum_{j \in In(v_i)} \frac{1}{|Out(v_j)|} S(v_j)$$

式(1)

(5)根据权重 $S(V_i)$ 对节点排序，获取得分最高的 T 个单词，作为该文本的关键词。

4.1.2 HDP 模型主题划分 Y. W. Teh 在 DP(Dirichlet Process，狄利克雷过程模型)基础上，于 2005 年提出 HDP(Hierarchical Dirichlet Process)层次狄利克雷过程模型[27]。相比目前主流的 LDA 主题模型，HDP 更具有处理强稀疏性数据、自动生成主题以及主题动态演化等优势。政府网站信息资源的单篇文档中通常反映了多个主题内容，倘若仅采用基于句子重要度或关键词的抽取方法，可能导致获取信息不完整，无法全面反映整篇文档的多个主题，为了抽取全领域知识元，需进行主题划分。在处理大数据文本时，HDP 模型能够挖掘文本深层语义且自动确定主题数目，抽取流程如图 3 所示：

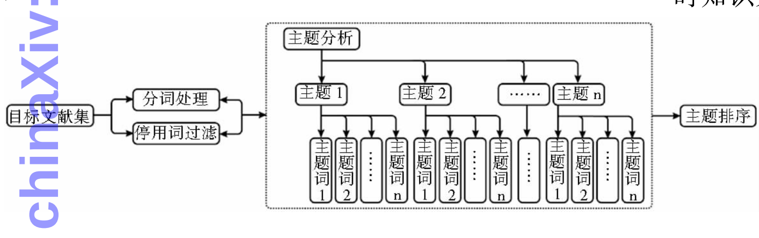


图 3　HDP 算法流程

(1)对目标文献集进行分词、停用词过滤、构建词频特征等预处理工作。

(2)对文献集合进行 HDP 主题分析，确定文献中 N 个主题的权重，依据抽取的具体主题中候选主题词权重，选择具体的主题词。其中主题 T 可以表示为：

$$T = (w_0, v_0; w_1, v_1; w_2, v_2; \cdots; w_n, v_n)$$

其中，T(topic)表示一个主题，主题 T 共有 n 个主题词，$w_n(word_n)$ 表示某一主题词，$v_n(value_n)$ 表示主题词在主题中的权重。

(3)依据贡献度进行主题排序。通过 TextRank 和 HDP 算法抽取关键词，经过人工对比分析，建立关键词集，根据政府网站信息资源知识元表示方法由领域专家确定知识元相关属性，并将完整的知识元存入知识元本体库。

4.2 政府网站信息资源领域知识元可视化表征模型

认知心理学认为信息的组织与外在表征直接影响人的认知、理解与内化[28]，可视化表征能够将承载知识的信息资源进行结构化组织，便于人们对知识形成整体且直观的认识，促进知识的获取、建构、应用和传播。知识可视化是利用视觉表征的手段构建与传达知识及知识间复杂的语义关系，可视化表征将有助于提升政府网站知识服务能力与知识应用价值。依据知识元模型理论与可视化思想，构建政府网站信息资源领域知识元可视化表征模型，该模型由领域知识元本体库生成模块与可视化知识服务模块构成。见图 4。

4.2.1 知识元本体库生成模块 政府网站信息资源知识元本体的构建是实现语义检索的基础，知识元本体用以描述资源本身以及资源间丰富的语义关系，使内容具有计算机理解的语义。利用上文提到的关键词与主题聚类方法抽取政府网站信息资源的元数据，人工对比筛选后组建关键词集，由领域专家选择本领域本体模板进行知识元标注，由于自动化标注技术的局限，目前主要采用半自动化的标注方法，依据标注结果获取知识元并建立临时知识元库，通过推理算法对临时知识元库进行动态推理，经人工识别修正不完善或不合理的知识元，将其存入知识元库。本体编辑器界面为用户提供语义标注上下文信息，允许和鼓励用户参与信息交流与共享，以获取和生成新的知识元，不断更新知识元库[1]。

4.2.2 可视化知识服务模块 政府网站知识可视化服务是建立在信息深度理解、分析与挖掘的基础上，为了提高用户知识检索的精准度，本文依据知识元模型理论、知识图谱思想和可视化技术对政府网站信息资源领域知识元进行可视化。在进行可视化之前需展开包括算法选择、阀值设置、时区分割、网络布局等准备工作，其中，算法选择可采用聚类算法(如 K-means、谱聚类算法等)、关联强度(如 BTM、HD、Jaccard 等)、特征词权重(如 TF. IDF、MI 等)；时区分割可依据分析年限节点进行时间切分；网络布局可采用 visio、2D/3DFruchterman-Reingold 、MDS、Kamada-Kawai 等算法计算语义相似的知识元在共现网络中的距离，帮助用户理解知识元库的知识内容。选用上述算法及功能进行自动标签，生成各类可视化知识图谱。

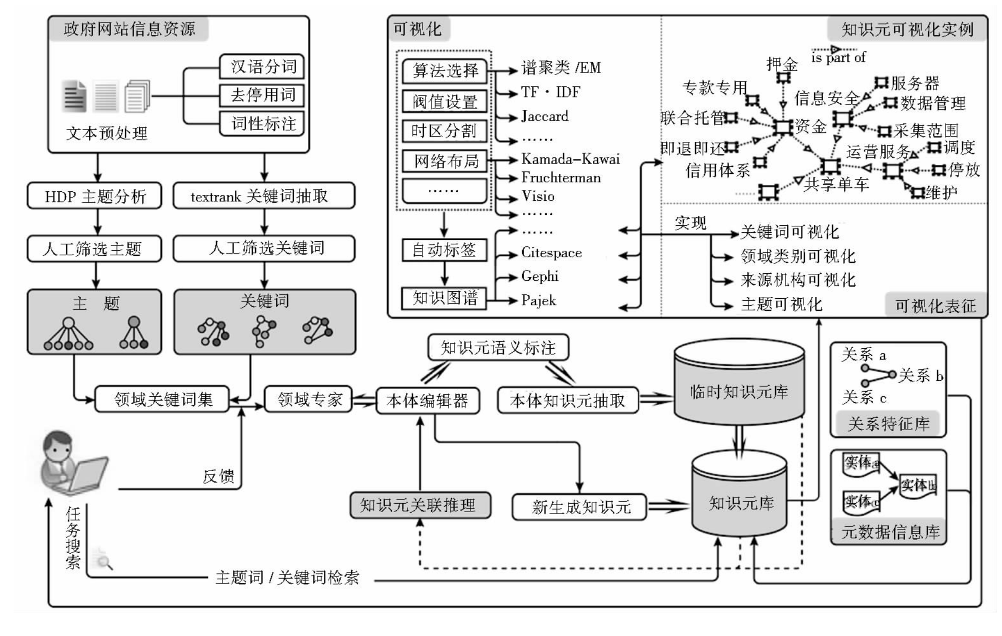


图 4 政府网站信息资源领域知识元可视化表征模型

用户在检索时可采用主题词或关键词的形式发出查询命令，政府网站信息系统将请求分解成一个查询指令交给知识元库，根据搜索任务匹配相应的知识元，借助上述相关算法及可视化工具可生成体现关键词共现、信息来源机构、政务信息主题与领域类别的可视化知识图谱[29]，用户根据查询结果向网站反馈服务满意度。

5 实例分析

共享单车作为移动互联网环境下共享经济创新模式的典型代表，在解决交通拥堵、短距离公共交通接驳等方面发挥着积极的作用，据《2017 年共享单车白皮书》报告显示，目前共享单车已遍布超过 20 个省、自治区的 50 多个城市；极光大数据也显示：截止到 2017 年 12 月，共享单车两大巨头 ofo 小黄车与摩拜单车的用户数量已分别达到 2 693.0 万人和 2 378.2 万人[30]。然而，共享单车在推动城市环境、交通、经济建设发展的同时，安全隐患、事故责任认定、资金管理、单车停放、市场运营秩序、单车技术性等问题也层出不穷，仅上海市 2017 年各类单车相关投诉案件就高达 7 978 件[31]。规范共享单车有序经营与发展已成为政府与公众广泛关注的热点问题，为此本文采集各级政府网站以“共享单车”“互联网共享自行车”相关主题的内容作为实验数据。

5.1 实验数据选取与预处理

为了确保数据的全面性与准确性，数据全部选取域名为“gov. cn”的网址，分别使用“共享单车 site：gov. cn”和“互联网共享自行车 site：gov. cn”作为关键词，通过百度进行相关文献的检索，时间截至 2018 年 3 月 20 日，共得到 1 131 个词条，利用 python 软件爬取这些词条对应的网址，并对这些网址进行过滤，其中网址中不包含 htm、asp 和 php 的网页一般为动态网页，不包含有用的正文，因此过滤掉网址不包含这 3 个关键词的网页。获取的网页结构包括导航栏、正文部分、评论部分等多个组成部分，在实际分析中采用哈尔滨工业大学开发的基于行块分布函数的通用网页正文抽取算法[32]，只抽取网页的正文部分。除此之外，由于防火墙和网页失效，部分网站无法爬取数据，在分析过程中不再关注这部分网页，过滤后共获取 797 个网页。在文本预处理阶段，采用 jieba 分词工具处理抽取的正文，为了提高文本挖掘的准确性，利用哈尔滨工业大学信息检索实验室开发的通用词表过滤停用词。

5.2 实验过程

5.2.1 关键词与主题抽取 采用 TextRank 算法抽取共享单车 797 篇相关文献的关键词，计算每个词在共现图中的重要程度，按排序结果选取前 30 个作为领域关键词，由于计算机抽取结果中包含部分冗余和无用词语，在自动抽取结果的基础上进行人工修正（去除区域、量词等词汇），关键词统计是基于出现这个关键词的文献数，即如果一个文献中存在这个关键词，无论出现多少次，只对该关键词统计一次，根据统计结果绘制图 5 与图 6：

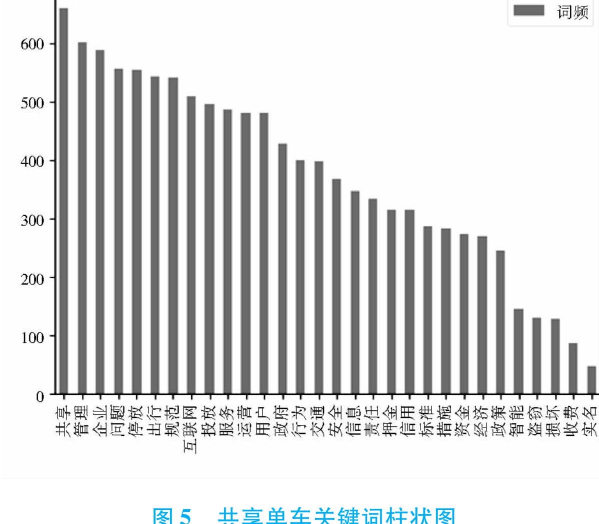


图5　共享单车关键词柱状图



图6　共享单车关键词云

为了获得更全面且准确的领域知识,除采用 TextRank 算法抽取关键词外,利用 HDP 主题模型算法再次对共享单车文本语料库进行主题建模,由于 HDP 不需要指定主题数且完全自动从语料库中学习主题,很大程度上可以提高大数据政府文献主题抽取的精准度。使用开源主题分析工具包 gensim 分析共享单车语料,抽取的主题以主题词(关键词)的形式表示,从中提取强度与贡献度最高的 20 个主题及主题词,经人工识别确定 3 个主题下的 30 个主题词,结果见表 2。

对比分析共享单车关键词与主题抽取结果,关键词抽取结果基本涵盖了代表该领域主题下的主题词,最终确定共享单车公共服务、运维监管、治理措施 3 个主题,以及该领域的 30 个关键词:共享、管理、企业、问题、停放、出行、规范、互联网、投放、服务、运营、用户、政府、行为、交通、安全、信息、责任、押金、信用、标准、措施、资金、经济、政策、智能、盗窃、损坏、收费、实名。

5.2.2　知识元"内容描述"属性抽取　通过对政府网站信息资源预处理、关键词抽取与主题识别,获取领域知识相关主题与关键词,由于计算机在领域知识分类与识别中存在一定偏差,在抽取中需进行人工干预,并

表2　共享单车主题模型抽取结果

公共服务	运维监管	治理措施
交通	停放	政府
低碳	规范	标准
经济	服务	责任
共享	运营	资金
接驳	投放	企业
绿色	二维码	信用
短距离	调度	安全
互联网	维护	自行车道
市民	损坏	承载力
最后一公里	盗窃	联合监管

将抽取结果与知识元本体库匹配,保留并存储可用知识元。依据前文建立的知识元六元组模型可知,知识元属性 D(内容描述)的抽取是获取其他知识元属性的基础,因此本文在实例研究部分重点探讨 D 的抽取工作以及介绍其他知识元属性抽取思路。以共享单车文献中"资金"为例,构建知识元属性:

I(索引号):信息标识码。

T(名称):该知识元的名称,即资金。

K(关键词):根据领域语料库抽取关键词,经人工筛选与识别,选取高频词作为该知识元的关键词集。资金知识元的关键词集为:押金、专款专用、联合托管、信用体系。

D(内容描述):鼓励免押金、建立信用体系;即租即压、即退即还;建立专用账户、专款专用、由金融机构联合监管。

C(类别):知识类别属性的抽取工作相对复杂,要结合领域本体知识的划分,本文中的共享单车属于交通范畴,而其中的资金是其子范畴,同时资金管理问题也可被视为金融相关领域范畴。这也是本文后续研究的重点工作。

O(地址):知识地址属性是知识元所在来源的超级链接,由于政府文献主题分布的特点,单篇文献涉及多个主题内容,实现该属性的抽取工作要在知识元内容属性基础上去冗余并聚类。

5.3　共享单车领域知识元可视化

依据共享单车知识元内容属性的抽取原理,结合政府网站信息资源领域知识元可视化表征模型,构建共享单车领域知识元可视化导航图,见图 7。

共享单车知识元可视化导航图以不同颜色和大小的节点及表达节点语义关系的连线来清晰地展示该领域的知识内容和知识结构。采用知识元为节点和知识元语义关系为连接线的层次结构化表达方式便于用户

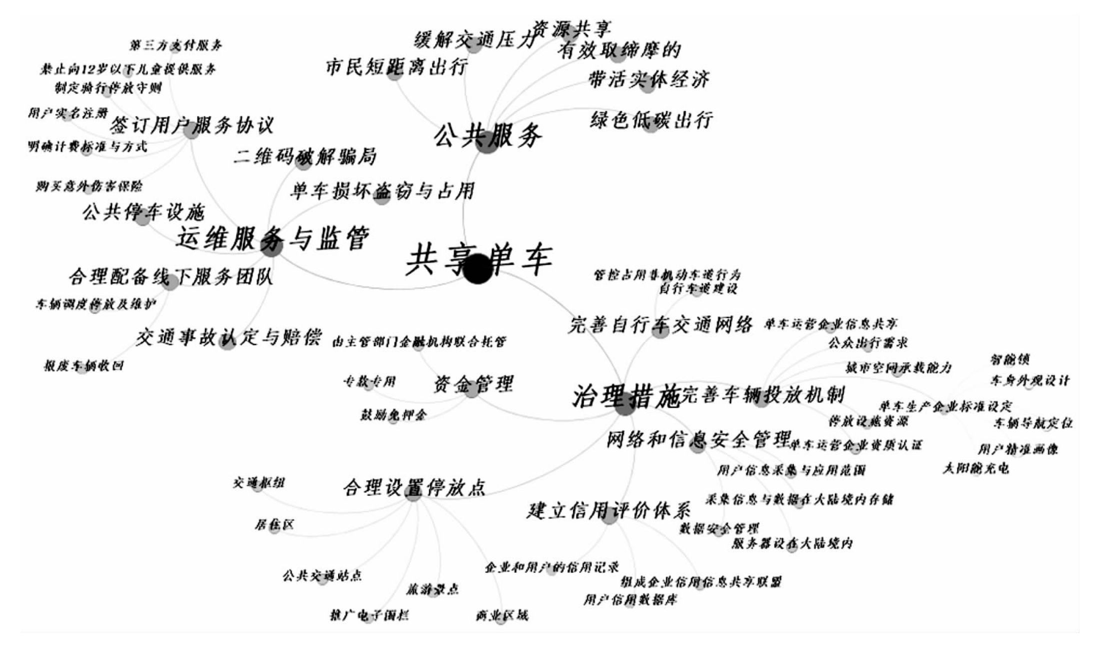


图 7　共享单车领域知识元可视化导航图

直观准确地做出检索与浏览的决策，只需点击搜寻任务的目标知识节点便可一键式或一站式浏览相符的知识内容。政府网站信息资源知识元可视化实现知识发现、知识整合、知识传播等知识服务满足用户认知的最大化需求，领域知识元结构化导航符合用户操作的最小努力原则，提升用户检索精度，有效促进信息交互效率。

6　结语

政府网站信息资源的多源异构和动态生长特性与政府网站网页文档粗粒度信息服务模式均不利于政务用户高效精确地获取和解读信息，因此本文从知识最小单位知识元出发研究政府网站信息资源细粒度知识服务范式。在知识元模型理论和知识可视化思想结合下构建了政府网站信息资源知识元可视化表征模型，通过共享单车实例验证了模型的可行性，表明该模型有助于提高政府网站信息资源的知识流转效率、知识发现概率以及知识服务精确率，为政府网站知识组织与知识服务提供新的研究路径。后续将根据政府网站信息资源知识元可视化表征模型设计开发知识元可视化原型系统。

参考文献：

[1] 温有奎，焦玉英. 基于知识元的知识发现[M]. 西安：西安电子科技大学出版社，2011.

[2] CNNIC 发布第 41 次《中国互联网络发展状况统计报告》[EB/OL]. [2018 – 03 – 01]. http://www.cac.gov.cn/2018-01/31/c_1122346138.htm.

[3] BOUGUETTAYA A，ELMAGARMID A，MEDJAHED B，et al. Ontology-based support for digital government[C]//Proceedings of 27th international conference on very large data bases. San Francisco：Morgan Kaufmann Publisher，2001：622 – 636.

[4] PROKOPIADOU G，PAPATHEODOROU C，MOSCHOPOULOS D. Integrating knowledge management tools for government information[J]. Government information quarterly，2004(2)：170 – 198.

[5] 高洁，李佳培. 电子政务信息资源管理中的领域本体构建研究[J]. 图书情报工作，2008，52(4)：57 – 61.

[6] 耿瑞利. 政府信息领域本体的构建研究[J]. 图书馆学研究，2010(11)：65 – 69.

[7] 邓峰，刘延滨. 基于云计算的政务信息融合平台的研究[J]. 情报杂志 2013，32(9)：123 – 127.

[8] 吕元智. 基于云计算的电子政务信息资源共享系统建设研究[J]. 情报理论与实践，2010，3(4)：106 – 109.

[9] 张玉涛，夏立新. 基于主题图的电子政务信息资源整合模型研究[J]. 情报杂志，2009，28(7)：161 – 165.

[10] 吕元智. 基于关联数据的电子政务信息资源语义组织研究[J]. 图书情报工作，2012，56(21)：143 – 146，130.

[11] 温有奎. 基于"知识元"的知识组织与检索[J]. 计算机工程与应用，2005(1)：55 – 57，91.

[12] 姜永常，杨宏岩，张丽波. 基于知识元的知识组织及其系统服务功能研究[J]. 情报理论与实践，2007，30(1)：37 – 40.

[13] 王宇，李秀秀. 基于知识元的期刊文献知识组织研究[C]//2017 第二届社会科学与人文国际会议论文集. 西安：科学出版集团，2017：9 – 14.

[14] 毕崇武，王忠义，宋红文. 基于知识元的数字图书馆多粒度集成知识服务[J]. 图书情报工作，2017，61(4)：115 – 122.

[15] SUN L. Knowledge element-based competitive intelligence analytics serving for SWOT situation assessment[C]// 8th international sym-

posium on computational intelligence and design. Hangzhou:IEEE, 2016:576－579.
[16] 于海峰,王延章,卢小丽,等.基于知识元的突发事件风险熵预测模型研究[J].系统工程学报,2016,31(1):117－126.
[17] 罗贤春,文庭孝,余肖生,等.社会化服务导向的电子政务信息资源共享目录[J].图书情报知识,2009(5):83－89.
[18] 马费成,赖茂生.信息资源管理[M].北京:高等教育出版社,2014.
[19] 周宁,余肖生,刘玮玮,等.基于XML平台的知识元表示与抽取研究[J].中国图书馆学报,2006(3):41－45.
[20] 温有奎,徐端颐,潘龙法.基于XML平台的知识元本体推理[J].情报学报,2004,23(6):643－648.
[21] 高国伟,王亚杰,李永先.知识元表示方法研究[J].现代情报,2015,35(3):15－18,33.
[22] 毕经元,顾新建,吕艳,等.基于知识元链接的汽车零部件知识管理系统[J].浙江大学学报,2009,43(12):2208－2212.
[23] 余洋.组织知识管理中的知识超网络研究[D].大连:大连理工大学,2009.
[24] 国务院办公厅秘书局关于印发政府信息公开目录系统实施指引(试行)的通知[EB/OL].[2018－08－05].http://www.gzw.dl.gov.cn/news/view_60146.html.
[25] YAO Y,ZENG Y, ZHONG N, et al. Knowledge retrieval(KR)[C]//Proceedings of IEEE/WIC/ACM International conference on Web intelligence. Silicon Valley:IEEE, 2007:729－735.
[26] MIHALCEA R,TARAU P. TextRank: bringing order into texts[C]//Proceedings of the empirical methods in natural language processing. Barcelona: Association for Computational Linguistics, 2004:404－411.
[27] TEH Y W, JORDAN M I,BEAL M J,et al. Hierarchical dirichlet processes[J]. Publications of the American Statistical Association, 2006,101(476):1566－1581.
[28] 斯腾伯格.认知心理学[M].邵志芳,译.北京:中国轻工业出版社, 2016.
[29] 李洁,毕强.数字图书馆资源知识聚合可视化模型构建研究[J].情报学报,2016,35(12):1273－1284.
[30] 大数据:头部共享单车app2017年12月运营报告[EB/OL].[2018－03－01].http://www.Jiemian.com/article/1905171.html.
[31] 共享单车成消费者投诉新热点,2017年上海收到投诉7978件[EB/OL].[2018－03－01].http://sh.qihoo.com/pc/2s182g1ger? sign=360_e39369d1.
[32] 基于行块分布函数的通用网页正文抽取[EB/OL].[2018－03－20].http://www.docin.com/p-656052252.html.

作者贡献说明：

王萍:负责论文修改及框架调整;

王美月:负责设计研究思路,收集与分析数据,撰写论文;

王益成:负责搜集文献,修改论文;

黄新平:负责论文修改及框架调整。

Study on Knowledge Element Model and Visual Representation of Government Website Information Resources

Wang Ping[1] Wang Meiyue[1] Wang Yicheng[1] Huang Xinping[2]
[1] School of Management, Jinlin University, Changchun 130022
[2] School of Public Policy & Management, Tsinghua University, Beijing 100084

Abstract: [**Purpose/significance**] The theory of knowledge element model is used to study the optimization path of the knowledge service effect of government Website, and the visual representation technology is helpful to reduce the operating load and the cognitive load of the information processing of the government users under the big data environment. [**Method/process**] According to the related knowledge element model, the six-tuple knowledge element representation method and the knowledge element ontology four tuple structures are deduced, which conforms to the characteristics of the information resources of the government Website. The TextRank and HDP algorithms are used to extract the key words and the subject words of the government Website information resources, and the domain experts determine the knowledge according to the extraction results. A visual representation model of government Website information resources knowledge element is constructed, which includes knowledge element ontology database generation and visual knowledge service. [**Result/conclusion**] The shared bicycle as an example issued by the government Website tests the effectiveness and feasibility of the visual representation model of knowledge element, and it provides a new research idea for the transition from government Website document's coarse-grained service to the knowledge element as a unit of fine-grained service, also with the help of visual knowledge services, the structured navigation of government information and the effect of user interpretation of domain text semantics are enhanced.

Keywords: government Website information resource　knowledge element　ontology　visual representation